PCT/AU 00/00056

09/890463

4

Patent Office
Canberra

AU00/56

I, KAY WARD, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PP 8463 for a patent by THE UNIVERSITY OF SYDNEY filed on 02 February 1999.

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

WITNESS my hand this
Seventh day of March 2000

K Ward

KAY WARD
TEAM LEADER EXAMINATION
SUPPORT AND SALES

AUSTRALIA

Patents Act 1990

THE UNIVERSITY OF SYDNEY

PROVISIONAL SPECIFICATION

Invention Title:

Pigment protein from coral tissue

The invention is described in the following statement:

Technical Field

The present invention relates to pigment proteins derived from corals, polynucleotide molecules encoding the pigment proteins, and uses thereof.

Background Art

There are a number of pigmented and/or fluorescent molecules which have been isolated and characterised from natural sources. Examples include apoaequorin, a single polypeptide chain isolated from the luminous jellyfish *Aequorea victoria*, green fluorescent protein (GFP), isolated from the luminous jellyfish *Aequorea victoria,* and renilla luciferase, isolated from the Renilla (also called sea pansies) which belong to a class of coelenterates known as anthozoans.

The present inventors have now purified pigment protein from coral tissue (PPCT) and cloned genes encoding PPCT which have unique and useful properties.

Disclosure of Invention

In a first aspect, the present invention consists in an isolated polynucleotide encoding a protein having the functional characteristics of pigment protein from coral tissue (PPCT).

Preferably, the isolated polynucleotide includes a nucleotide sequence selected from:

(i) a sequence of nucleotides shown in Figure 3 or 4;

(ii) a sequence of nucleotides shown in Figure 3 from nucleotide 1 to nucleotide 705;

(iii) a sequence of nucleotides shown in Figure 4 from nucleotide 1 to nucleotide 693;

(iv) a sequence capable of selectively hybridising to (i), (ii) or (iii) under stringent conditions; or

(v) a sequence encoding a polypeptide as defined below for the second aspect of the present invention.

Preferably, the isolated polynucleotide has a nucleotide sequence as shown in Figure 3 or 4.

In a preferred form, the isolated polynucleotide molecule encoding PPCT has at least 90%, more preferably 95%, identity to the nucleotide sequence shown in Figure 3, especially from nucleotide 1 to 705.

In an other preferred form, the isolated polynucleotide molecule encoding PPCT has at least 90%, more preferably 95%, identity to the nucleotide sequence shown in Figure 4, especially from nucleotide 1 to 693.

In a further preferred embodiment of the first aspect of the present invention, the polynucleotide sequence capable of hybridising to (i), (ii), or (iii) is less than 5000 nucleotides, however, it can be less than 1000 or even 500 nucleotides in length. Preferably, the hybridising polynucleotides of the present invention are at least 18 nucleotides in length.

A polynucleotide sequence of the present invention may hybridise to the sequence set out in Figure 3 or 4 under high stringency. As used herein, stringent conditions are those that (1) employ low ionic strength and high temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% $NaDodSO_4$ at 50^0C; (2) employ during hybridisation a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42^0C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 g/ml), 0.1% SDS and 10% dextran sulfate at 42^0C in 0.2 x SSC and 0.1% SDS

In a second aspect, the present invention consists in an isolated polypeptide having the functional characteristics of pigment protein from coral tissue (PPCT).

Preferably, the isolated polypeptide, includes an amino acid sequence selected from:

(i) a sequence of amino acids shown in Figure 5 or 6;

(ii) a unique sequence of amino acids shown in Figure 5 from amino acid 1 to amino acid 235;

(iii) a unique sequence of amino acids shown in Figure 6 from amino acid 1 to amino acid 231; or

(iv) a polypeptide encoded by an isolated polynucleotide as defined above for the first aspect of the present invention.

The polypeptide according to the present invention (PPCT) has the following characteristics:

i) Aqueous soluble protein (231-235 amino acids in length)

ii) Absorbance spectra with λ_{max} ranging from 555—600 nm depending on species

iii) Light induced

iv) Mostly in the form of a dimer, through sometimes trimeric

v) BlastX search reveals 22% amino acid identity with green fluorescent protein (GFP) and 56% amino acid similarity

vi) "QYG" in PPCT instead of "SYG" in GFP chromatophore position

vii) Estimated PI = 9.5

viii) Not fluorescent under normal conditions, through fluoresces at long wavelengths (above 560 nm) when corals are under stress (ie. pre-bleaching, potentially due to the presence of superoxides). GFP and its mutants fluoresces at wavelengths shorter than 540 nm.

It will be appreciated that the present invention includes modified and active forms of PPCT.

PPCT can be obtained from the tissues of the following coral families: Pocilloporidae, Acroporidae, Poritidae, Merulinidae, Fungiidae.

Preferably, PPCT is obtained from the tissue from the coral: *Acropora aspera*.

In a third aspect, the present invention consists in a suitable vector for the replication and/or expression of a polynucleotide according to the first aspect of the present invention. The vectors may be, for example, plasmid, virus or phage vectors provided with an origin of replication, and preferably a promoter for the expression of the polynucleotide and optionally a regulator of the promoter. The vector may contain one or more selectable markers, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian expression vector. The vector may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell.

In a fourth aspect of the present invention relates to host cells transformed or transfected with the vector of the third aspect of the present invention.

Suitable host cells for cloning or expressing the protein(s) disclosed herein are the prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include eubacteria, such as Gram-negative or Gram-positive organisms, for example, *E. coli*, *Bacilli* such as *B. subtilis* or *B. thuringiensis*,

Pseudomonas species such as *P. aeruginosa*, *Salmonella typhimurium* or *Serratia marcescens*.

Eukaryotic microbes such as filamentous fungi or yeast are suitable hosts for expressing the protein(s) of the present invention. *Saccharomyces cerevisiae*, or common baker's yeast, is the most commonly used among lower eukaryotic host microorganisms. However, a number of other genera, species, and strains are commonly available and useful herein, such as *Schizosaccharomyces pombe*; *Kluyveromyces* hosts such as e.g. *K. lactis*; filamentous fungi such as, e.g. *Neurospora*, or *Penicillium*; and *Aspergillus* hosts such as *A. nidulans* and *A. niger*.

Suitable higher eukaryotic host cells can be cultured vertebrate, invertebrate or plant cells. Insect host cells from species such as *Spodoptera frugiperda*, *Aedes aegypti*, *Aedes albopictus*, *Drosophila melanogaster*, and *Bombyx mori* can be used. Plant cell cultures of cotton, corn, potato, soybean, tomato, and tobacco can be utilised as hosts. Typically, plant cells are transfected by incubation with certain strains for the bacterium *Agrobacterium tumefaciens*.

Propagation of vertebrate cells in culture (tissue culture) has become a routine procedure in recent years. Examples of useful mammalian host cell lines are monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture); baby hamster kidney cells (BHK ATCC CCL 10); Chinese hamster ovary cells/-DHFR (CHO); mouse sertoli cells, monkey kidney cells (CV1 ATCC CCL 70); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK ATCC CCL 34), and a human hepatoma cell line (Hep G2). Preferred host cells are human embryonic kidney 293 and Chinese hamster ovary cells.

Host cells are transfected and preferably transformed with expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. Transformation means introducing DNA into an organism so that the DNA is replicable, either as an extrachromosomal element or by chromosomal integration. Depending on the host cell used, transformation is done using standard techniques appropriate to such cells.

In a fifth aspect, the present invention consist in a process for preparing a polypeptide according to the second aspect of the present invention, the process includes cultivating a host cell transformed or transfected with an (expression) vector of the third aspect of the present invention under conditions providing for expression of the polynucleotide encoding the polypeptide, and recovering the expressed polypeptide. Such cells can be used for the production of commercially useful quantities of the encoded protein.

In an sixth aspect, the present invention consists in an oligonucleotide probe or primer, the probe or primer having a sequence that hybridises selectively to a polynucleotide according to the first aspect of the present invention.

In a preferred embodiment of the sixth aspect, the oligonucleotide probe or primer includes at least 8 nucleotides, more preferably at least 18 nucleotides and more preferably at least 25 nucleotides.

In a further preferred embodiment, the oligonucleotide probe or primer is used as a detectable probe where the oligonucleotide is conjugated with a label such as a radioisotope, an enzyme, biotin, a fluorescent molecule or a chemiluminescent molecule.

In a seventh aspect, the present invention consists in the use of PPCT as a tissue marker, fluorescent marker, and general dye.

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

In order that the present invention may be more clearly understood, preferred forms will be described with reference to the following examples and drawings.

Brief Description of Drawings

Figure 1 shows N-terminal sequences of PPCT from a number of coral sources.

Figure 2 shows a nucleotide sequence of a degenerative primer useful for hybridising to genes encoding PPCT.

Figure 3 shows a cDNA nucleotide sequence of clone T7SP6BASPOC4 encoding PPCT.

Figure 4 shows a cDNA nucleotide sequence of clone T7SP6BASPOC3 encoding PPCT.

Figure 5 shows a amino acid sequence encoded by clone T7SP6BASPOC4 of Figure 3.

Figure 6 shows a amino acid sequence encoded by clone T7SP6BASPOC3 of Figure 4.

Figure 7 shows the enzymes that do not cut the polynucleotide encoding PPCT.

Figure 8 shows a restriction enzyme map of the cDNA encoding PPCT.

Figure 9 shows a vector for including cDNA nucleotide sequence encoding PPCT.

Modes for Carrying Out the Invention

General Molecular Biology

Unless otherwise indicated, the recombinant DNA techniques utilised in the present invention are standard procedures, well known to those skilled in the art. Such techniques are described and explained throughout the literature in sources such as, J. Perbal, A Practical Guide to Molecular Cloning, John Wiley and Sons (1984), J. Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory Press (1989), T.A. Brown (editor), Essential Molecular Biology: A Practical Approach, Volumes 1 and 2, IRL Press (1991), D.M. Glover and B.D. Hames (editors), DNA Cloning: A Practical Approach, Volumes 1-4, IRL Press (1995 and 1996), and F.M. Ausubel et al. (Editors), Current Protocols in Molecular Biology, Greene Pub. Associates and Wiley-Interscience (1988, including all updates until present) and are incorporated herein by reference.

Gene/DNA Isolation

The DNA encoding a protein may be obtained from any cDNA library prepared from tissue believed to express the gene mRNA and to express it at a detectable level. DNA can also be obtained from a genomic library.

Libraries are screened with probes or analytical tools designed to identify the gene of interest or the protein encoded by it. For cDNA expression libraries, suitable probes include monoclonal or polyclonal antibodies that recognise and specifically bind the protein; oligonucleotides of about 20-80 bases in length that encode known or suspected portions of cDNA from the same or different species; and/or complementary or homologous cDNAs or fragments thereof that encode the same or a

hybridising gene. Appropriate probes for screening genomic DNA libraries include, but are not limited to, oligonucleotides; cDNAs or fragments thereof that encode the same or hybridising DNA including expressed sequence tags and the like; and/or homologous genomic DNAs or fragments thereof. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures as described in chapters 10-12 of Sambrook *et al.*

An alternative means to isolate a gene encoding is to use polymerase chain reaction (PCR) methodology as described in section 14 of Sambrook *et al.* This method requires the use of oligonucleotide probes that will hybridise to the gene.

The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimised. The actual nucleotide sequence(s) is usually based on conserved or highly homologous nucleotide sequences or regions of the gene. The oligonucleotides may be degenerate at one or more positions. The use of degenerate oligonucleotides may be of particular importance where a library is screened from a species in which preferential codon usage in that species is known. The oligonucleotide must be labelled such that it can be detected upon hybridisation to DNA in the library being screened. The preferred method of labelling is to use ^{32}P-labelled ATP with polynucleotide kinase, as is well known in the art, to radiolabel the oligonucleotide. However, other methods may be used to label the oligonucleotide, including, but not limited to, biotinylation or enzyme labelling.

Nucleic acid having all the protein coding sequence is obtained by screening selected cDNA or genomic libraries, and if necessary, using conventional primer extension procedures as described in section 7.79 of Sambrook *et al.*, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

Another alternative method for obtaining the gene of interest is to chemically synthesise it using one of the methods described in Fingels *et al.* (Agnew Chem. Int. Ed. Engl. *28:* 716-734, 1989). These methods include triester, phosphite, phosphoramidite and H-Phosphonate methods, PCR and other autoprimer methods, and oligonucleotide syntheses on solid supports. These methods may be used if the entire nucleic acid sequence of the gene is known, or the sequence of the nucleic acid complementary to the coding

strand is available, or alternatively, if the target amino acid sequence is known, one may infer potential nucleic acid sequences using known and preferred coding residues for each amino acid residue.

Mutants, Variants and Homology - Nucleic Acids

Mutant polynucleotides will possess one or more mutations which are deletions, insertions, or substitutions of nucleotide residues. Mutants can be either naturally occurring (that is to say, isolated from a natural source) or synthetic (for example, by performing site-directed mutagensis on the DNA). It is thus apparent that polynucleotides of the invention can be either naturally occurring or recombinant (that is to say prepared using recombinant DNA techniques).

An allelic variant will be a variant that is naturally occurring within an individual organism.

Nucleotide sequences are homologous if they are related by divergence from a common ancestor. Consequently, a species homologue of the polynucleotide will be the equivalent polynucleotide which occurs naturally in another species. Within any one species a homologue may exist as numerous allelic variants, and these will be considered homologues of the polynucleotide. Allelic variants and species homologues can be obtained by following standard techniques known to those skilled in the art. Preferred species homologues include those obtained from representatives of the same Phylum, more preferably the same Class and even more preferably the same Order.

A polynucleotide at least 70% identical, as determined by methods well known to those skilled in the art (for example, the method described by Smith, T.F. and Waterman, M.S. (1981) Ad. Appl. Math., 2: 482-489, or Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol., 48: 443-453), to the that of the present invention are included in the invention, as are proteins at least 80% or 90% and more preferably at least 95% identical to the polynucleotide of the present invention. This will generally be over a region of at least 60, preferably at least 90, contiguous nucleotide residues.

Mutants, Variants and Homology - Proteins

Mutant polypeptides will possess one or more mutations which are deletions, insertions, or substitutions of amino acid residues. Mutants can be either naturally occurring (that is to say, purified or isolated from a natural source) or synthetic (for example, by performing site-directed mutagensis on

the encoding DNA). It is thus apparent that polypeptides of the invention can be either naturally occurring or recombinant (that is to say prepared using recombinant DNA techniques).

An allelic variant will be a variant that is naturally occurring within an individual organism.

Protein sequences are homologous if they are related by divergence from a common ancestor. Consequently, a species homologue of the protein will be the equivalent protein which occurs naturally in another species. Within any one species a homologue may exist as numerous allelic variants, and these will be considered homologues of the protein. Allelic variants and species homologues can be obtained by following standard techniques known to those skilled in the art. Preferred species homologues include those obtained from representatives of the same Phylum, more preferably the same Class and even more preferably the same Order.

A protein at least 50% identical, as determined by methods well known to those skilled in the art (for example, the method described by Smith, T.F. and Waterman, M.S. (1981) Ad. Appl. Math., 2: 482-489, or Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol., 48: 443-453), to the that of the present invention are included in the invention, as are proteins at least 70% or 80% and more preferably at least 90% identical to the protein of the present invention. This will generally be over a region of at least 20, preferably at least 30, contiguous amino acids.

Protein Variants

Amino acid sequence variants can be prepared by introducing appropriate nucleotide changes into DNA, or by *in vitro* synthesis of the desired polypeptide. Such variants include, for example, deletions, insertions or substitutions of residues within the amino acid sequence. A combination of deletion, insertion and substitution can be made to arrive at the final construct, provided that the final protein product possesses the desired characteristics. The amino acid changes also may alter post-translational processes such as changing the number or position of glycosylation sites, altering the membrane anchoring characteristics, altering the intra-cellular location by inserting, deleting or otherwise affecting the transmembrane sequence of the native protein, or modifying its susceptibility to proteolytic cleavage.

In designing amino acid sequence variants, the location of the mutation site and the nature of the mutation will depend on characteristic(s) to be modified. The sites for mutation can be modified individually or in series, e.g., by (1) substituting first with conservative amino acid choices and then with more radical selections depending upon the results achieved, (2) deleting the target residue, or (3) inserting residues of other ligands adjacent to the located site.

A useful method for identification of residues or regions for mutagenesis is called "alanine scanning mutagenesis" as described by Cunningham and Wells (*Science* (1989) **244**: 1081-1085*)*. Here, a residue or group of target residues are identified (e.g., charged residues such as Arg, Asp, His, Lys, and Glu) and replaced by a neutral or negatively charged amino acid (most preferably alanine or polyalanine) to affect the interaction of the amino acids with the surrounding aqueous environment in or outside the cell. Those domains demonstrating functional sensitivity to the substitutions then are refined by introducing further or other variants. Thus, while the site for introducing an amino acid sequence variation is predetermined, the nature of the mutation *per se* need not be predetermined. For example, to optimise the performance of a mutation at a given site, alanine scanning or random mutagenesis may be conducted at the target codon or region and the expressed variants are screened for the optimal combination of desired activity.

There are two principal variables in the construction of amino acid sequence variants: the location of the mutation site and the nature of the mutation. These may represent naturally occurring alleles or predetermined mutant forms made by mutating the DNA either to arrive at an allele or a variant not found in nature. In general, the location and nature of the mutation chosen will depend upon the characteristic to be modified.

Amino acid sequence deletions generally range from about 1 to 30 residues, more preferably about 1 to 10 residues and typically about 1 to 5 contiguous residues.

Amino acid sequence insertions include amino and/or carboxyl-terminal fusions ranging in length from one residue to polypeptides containing a hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Other insertional variants include the fusion of the N- or C-terminus of the proteins to an immunogenic

polypeptide e.g. bacterial polypeptides such as betalactamase or an enzyme encoded by the *E. coli trp* locus, or yeast protein, bovine serum albumin, and chemotactic polypeptides. C-terminal fusions with proteins having a long half-life such as immunoglobulin constant regions (or other immunoglobulin regions), albumin, or ferritin, are included.

Another group of variants are amino acid substitution variants. These variants have at least one amino acid residue in the protein molecule removed and a different residue inserted in its place. The sites of greatest interest for substitutional mutagenesis include sites identified as the active site(s). Other sites of interest are those in which particular residues obtained from various species are identical. These positions may be important for biological activity. These sites, especially those falling within a sequence of at least three other identically conserved sites, are substituted in a relatively conservative manner. Such conservative substitutions are shown in Table 1 under the heading of "preferred substitutions". If such substitutions result in a change in biological activity, then more substantial changes, denominated exemplary substitutions in Table 1, or as further described below in reference to amino acid classes, are introduced and the products screened.

TABLE 1 Preferred amino acid substitutions

Original Residue	**Exemplary Substitutions**	**Preferred Substitutions**
Ala (A)	val; leu; ile	val
Arg (R)	lys; gln; asn	lys
Asn (N)	gln; his; lys; arg	gln
Asp (D)	glu	glu
Cys (C)	ser	ser
Gln (Q)	asn	asn
Glu (E)	asp	asp
Gly (G)	pro	pro
His (H)	asn; gln; lys; arg	arg
Ile (I)	leu; val; met; ala; phe norleucine	leu
Leu (L)	norleucine, ile; val; met; ala; phe	ile
Lys (K)	arg; gln; asn	arg
Met (M)	leu; phe; ile;	leu
Phe (F)	leu; val; ile; ala	leu
Pro (P)	gly	gly
Ser (S)	thr	thr
Thr (T	ser	ser
Trp (W)	tyr	tyr
Tyr (Y)	trp; phe; thr; ser	phe
Val (V)	ile; leu; met; phe ala; norleucine	leu

Substantial modifications in function or immunological identity are accomplished by selecting substitutions that differ significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydro-phobicity of the molecule at the target site, or (c) the bulk of the side chain. Naturally occurring residues are divided into groups based on common side chain properties:

(1) hydrophobic: norleucine, met, ala, val, leu, ile;

(2) neutral hydrophilic: cys, ser, thr;

(3) acidic: asp, glu;

(4) basic: asn, gln, his, lys, arg;

(5) residues that influence chain orientation: gly, pro; and

(6) aromatic: trp, tyr, phe

Non-conservative substitutions will entail exchanging a member of one of these classes for another.

Substantially Purified

By "substantially purified" the present inventors mean a polypeptide that has been separated from lipids, nucleic acids, other polypeptides, and other contaminating molecules.

Active Fragment

By "active fragment" the present inventors mean a fragment of a sequence shown in Figure 5 or 6 which retains at least one of the activities of the native PPCT polypeptide.

RESULTS

Purification Scheme

(i) Extraction of coloured pigment from coral tissue by immersing coral overnight at 4°C in 0.6 M Potassium Phosphate buffer pH 6.65.

(ii) Gel filtration (Pharmacia, Superose HR 10/30) monitored at 280 nm & λ_{max} eluting with Phosphate buffer (peak collection)

(iii) Desalt and lyophilise

(iv) Separate on 15% SDS-PAGE

(v) Electroblot onto PVDF using CAPS buffer

(vi) N-terminal Sequence analysis (Australian National University Sequencing Facility)

cDNA Library

Species: Blue tipped *Acropora aspera*

(i) Isolation of mRNA: Qiagen nRNA Kit (poly A+)

(ii) cDNA made with SMART PCR cDNA Synthesis Kit from CLONTECH

(iii) cDNA ligated into Stratagene LambdaZapII EcoR1 cut/CIAP and Packaged by Stratagene Gigapack II.

PCR out of Library

(i) Used gel purified oligomer as 5' primers and SMART PCR adaptor 3' primers - PCR amplified genes

(ii) Gel purified product ligated into pGemT easy vector (Promega)

LICOR sequencing from T7 and SP6 primers

Translated sequence to give Amino acid sequence

Properties of Pigment Protein from Coral

i) Aqueous soluble protein (231-235 amino acids in length)

ii) Absorbance spectra with λ_{max} ranging from 560—590 nm depending on species

iii) Light induced

iv) Mostly in the form of a dimer, through sometimes trimeric

v) BlastX search reveals 22% amino acid identity with green fluorescent protein (GFP) and 56% amino acid similarity

vi) "QYG" in PPCT instead of "SYG" in GFP chromatophore position.

vii) Estimated PI = 9.5

viii) Not fluorescent under normal conditions, through fluoresces at long wavelengths (above 560 nm) when corals are under stress (ie. pre-bleaching, potentially due to the presence of superoxides). Note that GFP and its mutants fluoresces at wavelengths shorter than 540 nm.

Potential Uses

i) Tissue marker. Because it is blue or pink (depending on the species) it should have potential as a harmless protein based marker protein for following gene expression in transformed tissues.

ii) Mutants can be probably made to fluoresce. Because it absorbs light at longer wavelengths than GFP, it has a potential to fluoresce at longer wavelengths and with a greater variety of colours than GFP.

iii) PPCT also has potential as easily manufactured harmless dye. It is blue in Acroporids and Poritids, and pink in Pocilloporids. Because it can be synthesised in tranformed bacteria, it should be easily manufactured.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Dated this second day of February 1999

THE UNIVERSITY OF SYDNEY
Patent Attorneys for the Applicant:

F B RICE & CO

N-TERMINAL SEQUENCE OF COLORED TISSUE PROTEINS FROM:

Montipora caliculata
SVIAKQMTYKVYMSGTVN

λ_{max} = 574-578
subunit MW= 28 kD
native = dimer

Montipora monasteriata
SVIAK

λ_{max} = 579, shoulder at approx. 545 nm
subunit MW= 28 kD
native = dimer

Porites murrayensis
SVIAKQMTYKVYMSGTVN

λ_{max} = 576, shoulder at approx. 545 nm
subunit MW= 28 kD
native = dimer

Porites lobata
SVIAKQMTYKVYMSGTVN**NHYEFVT**
In bold, doubt over sequence.

λ_{max} = 576, shoulder at approx. 545 nm
subunit MW= 28 kD
native = dimer

Acropora horrida
SVIAKQMTYKVYMSGTV

λ_{max} = 579, shoulder at approx. 545 nm
subunit MW= 28 kD
native = dimer

Figure 1

DEGENERATIVE OLIGOMER (30-MER):

TCCGTTATCGCTAAACAGATGACCTACAAA

***Acropora aspera* derived sequence:**

SVIAKQMTYKVYMSGTVNGHYFEVE (1^{ST} 25AA)

In bold, differences from estimated *P. lobata* sequence

Figure 2

COLORED PROTEIN cDNA SEQUENCE FROM *A. aspera*

T7SP6BASPOC4 Length: 841 (today) Check: 7145 ..

```
   1  TCCGTTATCG CTAAACAGAT GACCTACAAA GTTTATATGT CAGGCACGGT
  51  CAATGGACAC TACTTTGAGG TCGAAGGCGA TGGAAAAGGA AAGCCTTACG
 101  AGGGGGAGCA GACGGTAAGG CTGGCTGTCA CCAAGGGCGG ACCTCTGCCA
 151  TTTGCTTGGG ATATTTTATC ACCACAGTGT CAGTACGGAA GCATACCATT
 201  CACCAAGTAC CCTGAAGACA TCCCTGACTA TGTAAAGCAG TCATTCCCGG
 251  GGAGATATAC ATGGGAGAGG ATCATGAACT TTGAAGATGG TGCAGTGTGT
 301  ACTGTCAGCA ATGATTCCAG CATCCAAGGC AACTGTTTCA TCTACCATGT
 351  CAAGTTCTCT GGTTTGAACT TTCCTCCCAA TGGACCTGTT ATGCAGAAGA
 401  AGACACAGGG CTGGGAACCC AACACTGAGC GTCTCTTTGC ACGAGATGGA
 451  ATGCTGATAG GAAACAACTT TATGGCTCTG AAGTTAGAAG GAGGTGGTCA
 501  CTATTTGTGT GAATTCAAAT CTACTTACAA GGCAAAGAAG CCTGTGAAGA
 551  TGCCAGGGTA TCACTATGTT GACCGCAAAC TGGATGTAAC CAATCACAAC
 601  AAGGATTACA CTTCCGTTGA GCAGTGTGAA ATTTCCATTG CACGCAAACC
 651  TGTGGTCGCC TGCCGTTTTT TCAGAGTCAA ATCAAGGCAC AAATACGCAG
 701  TGGCGTAAAA AACGTAGATT CTGATTTTAG CTTATAGAAG TAGGAACGAA
 751  GAAGTGTAAA CAACCATTAA TGATTAAACT TTTGAAAACA ACGCCATAAA
 801  AAAAAAAAAA AAAAAAAAAA AAAAAGCGGC CGCTCGAATT A
```

Figure 3

COLORED PROTEIN cDNA SEQUENCE FROM *A. aspera*

t7SP6BASPOC3 Length: 841 (today) Check: 6299 ..

```
   1  TCCGTTATCG CTAAACAGAT GACCTACAAA GTTTATATGT CAGGCACGGT
  51  CAATGGACAC TACTTTGAGG TCGAAGGCGA TGGAAAAGGA AAGCCTTACG
 101  AGGGGGAGCA GACGGTAAGG CTGGCTGTCA CCAAGGGCGG ACCTCTGCCA
 151  TTTGCTTGGG ATATTTTATC ACCACAGTGT CAGTACGGAA GCATACCATT
 201  CACCAAGTAC CCTGAAGACA TCCCTGACTA TGTAAAGCAG TCATTCCCGG
 251  GGAGATATAC ATGGGAGAGG ATCATGAACT TTGAAGATGG TGCAGTGTGT
 301  ACTGTCAGCA ATGATTCCAG CATCCAAGGC AACTGTTTCA TCTACCATGT
 351  CAAGTTCTCT GGTTTGAACT TTCCTCCCAA TGGACCTGTT ATGCAGAAGA
 401  AGACACAGGG CTGGGAACCC AACACTGAGC GTCTCTTTGC ACGAGATGGA
 451  ATGCTGATAG GAAACAACTT TATGGCTCTG AAGTTAGAAG GAGGTGGTCA
 501  CTATTTGTGT GAATTCAAAT CTACTTACAA GGCAAGGAAG CCTGTGAAGA
 551  TGCCAGGGTA TCACTATGTT GACCGCAAAC TGGATGTAAC CAATCACAAC
 601  AAGGATTACA CTTCCGTTGA GCAGCGTGAA ATTTCCATTG CACGCAAACC
 651  TTTGGTCGCC TGCTGTTTTT TCAGAGTCAA ATCAAGGCAC AAATAAGCAG
 701  TGGCGTAAAA AACGTAGATT CTGATTTTAG CTTAGAGAAG TAGGAACGAA
 751  GAAGTGTAGA CAACCTTCAA TGATTAAACT TTTGAAAACA ACSCCAAAAA
 801  AAAAAAAAAA AAAAAAAAAA AAAAAGCGGC CGCTCGAATT A
```

663 either A or C
In **bold** differences

Figure 4

TRANSLATED COLORED PROTEIN CDNA FROM A. *aspera*

BASPOC4.pep

```
   1  SVIAKQMTYK VYMSGTVNGH YFEVEGDGKG KPYEGEQTVR LAVTKGGPLP
  51  FAWDILSPQC QYGSIPFTKY PEDIPDYVKQ SFPGRYTWER IMNFEDGAVC
 101  TVSNDSSIQG NCFIYHVKFS GLNFPPNGPV MQKKTQGWEP NTERLFARDG
 151  MLIGNNFMAL KLEGGGHYLC EFKSTYKAKK PVKMPGYHYV DRKLDVTNHN
 201  KDYTSVEQCE ISIARKPVVA CRFFRVKSRH KYAVA*
```

Figure 5

TRANSLATED COLORED PROTEIN CDNA FROM *A. aspera*

BASPOC3.pep

```
      1  SVIAKQMTYK VYMSGTVNGH YFEVEGDGKG KPYEGEQTVR LAVTKGGPLP
     51  FAWDILSPQC QYGSIPFTKY PEDIPDYVKQ SFPGRYTWER IMNFEDGAVC
    101  TVSNDSSIQG NCFIYHVKFS GLNFPPNGPV MQKKTQGWEP NTERLFARDG
    151  MLIGNNFMAL KLEGGGHYLC EFKSTYKARK PVKMPGYHYV DRKLDVTNHN
    201 KDYTSVEQRE ISIARKPLVA CCFFRVKSRH K*
```

In **bold**, differences between clones.

Figure 6

Enzymes that do not cut:

AatI	ApaI	HindIII	NcoI	NdeI	NsiI	PstI
SacI	SacII	SalI	ScaI	SpeI	SphI	XbaI
XhoI	XhoII	XmnI	BstZ	NotI		

Figure 7

RESTRICTION ENZYME MAP OF COLORED PROTEIN cDNA

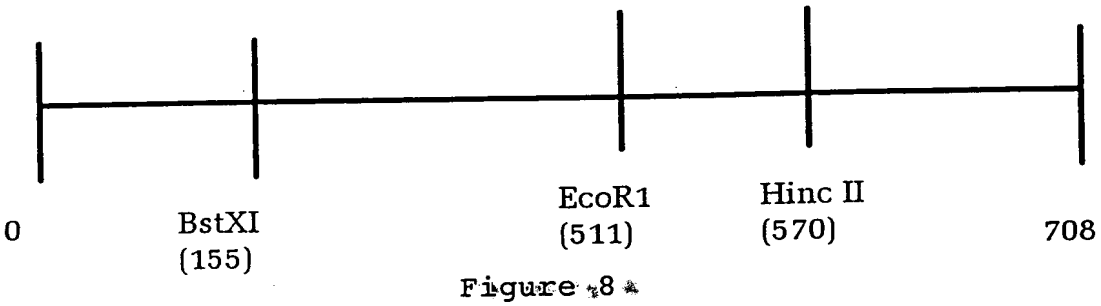

Figure 8

COLORED PROTEIN VECTOR

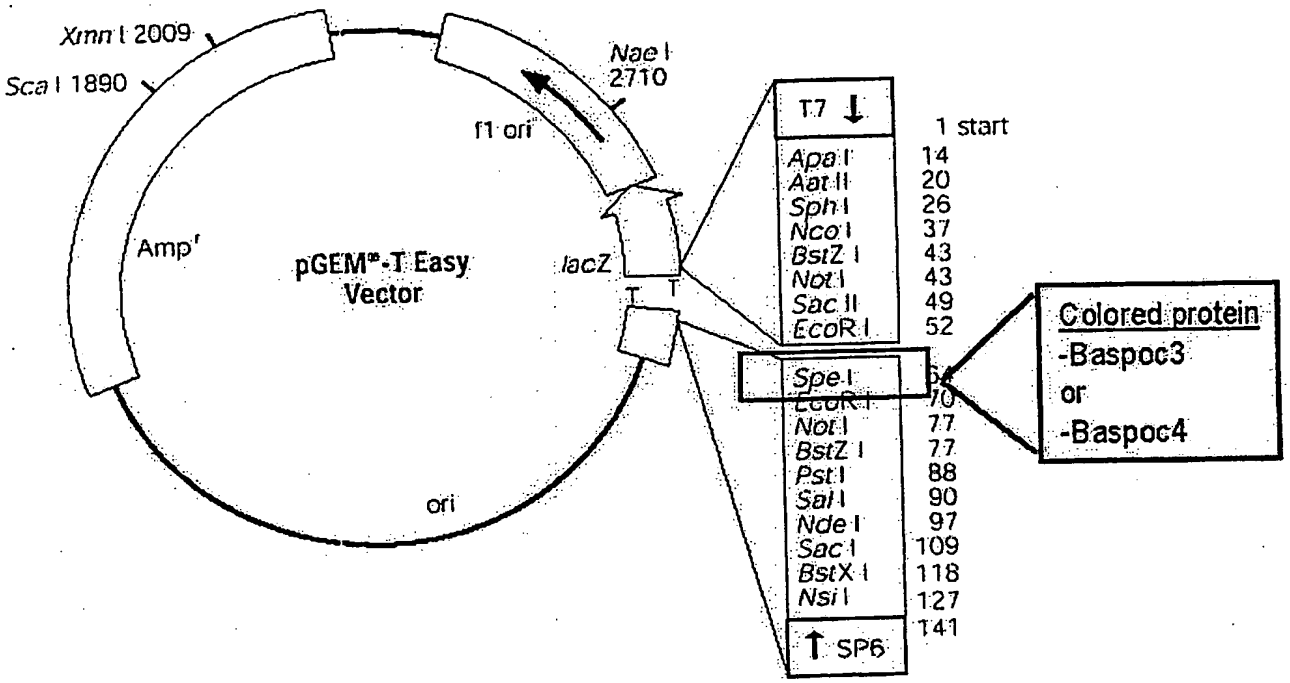

Figure 9